DOCUMENT RESUME

ED 266 742 HE 019 112

AUTHOR Kulik, James A.
TITLE Using Student Ratings in Evaluating Teaching Assistants.
PUB DATE Mar 85
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 31-April 4, 1985).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Oriented Programs; *Course Evaluation; Faculty Evaluation; *Feedback; Foreign Students; Graduate Students; Higher Education; *Student Evaluation of Teacher Performance; *Student Teacher Evaluation; *Teaching Assistants
IDENTIFIERS *University of Michigan

ABSTRACT

The use of a computer-based system of individualized course evaluations, Instructor-Designed Questionnaires (IDQs), to collect student reactions to classes at the University of Michigan is discussed. Originally designed to provide feedback to university teachers alone, the system is now used extensively by teachers to gather information for themselves as well as for departmental committees and student groups. The system is also used on a more limited basis in program evaluation. Recently, IDQs were used to collect student reactions to both foreign-educated and American teaching assistants (TAs) at the university. Initial ratings of both groups of teaching assistants were far below university norms, but ratings improved after one semester of teaching. It is suggested that a TA's first semester ratings are open to misinterpretation and that verbal facility is very important to teaching. IDQs (and similar course evaluation systems) were found to provide useful information for teachers wishing to improve their teaching, but the proportion of teachers who will use this type of system voluntarily is probably small. In addition to providing useful data for administrators and committees deliberating about personnel, these systems can also provide useful data for students choosing classes. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED266742

Using Student Ratings in Evaluating Teaching Assistants

James A. Kulik

Center for Research on Learning and Teaching

The University of Michigan

A Symposium Presentation at the Annual Meeting
Of the American Educational Research Association
Chicago, March 1985

HE019 112

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James A. Kulik

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

A computer-based system of individualized course evaluations--Instructor-Designed Questionnaires, or IDQ--has been used to collect student reactions to classes at the University of Michigan since 1975. Originally designed to provide feedback to University teachers alone, the system is now used extensively by teachers to gather information for themselves as well as for departmental committees and student groups. The system is also used on a more limited basis in program evaluation. Recently, for example, IDQs were used to collect student reactions to both foreign-educated and American teaching assistants at the University. Initial ratings of both groups of teaching assistants were far below University norms, but ratings improved after one semester of teaching.

My topic this afternoon is the University of Michigan's system of Instructor-Designed Questionnaire, or IDQ. Actually, I plan to talk about three systems. The first is the system that we originally designed--the drawing-board version of IDQ. The second is the system that faculty at Michigan currently use--the road-tested model. And the third is the system as it will be in the future as new evaluation needs emerge at Michigan.

Initial System

We first developed IDQ ten years ago as a way of making course evaluation forms more responsive to teacher concerns. Our system was patterned after the first computer-based, flexible system of course evaluation to be extensively used in American higher education--Purdue's Cafeteria system--and IDQ and Cafeteria shared many features. IDQ was meant to be distinct, however, in at least two ways. First, its materials were designed to prompt teachers to engage in self-assessment of their teaching at the outset of the evaluation process. The catalog of items used in the system was carefully constructed to guide instructors systematically through this self-assessment. Second, the system was meant to encourage colleague interactions on teaching. An important part of the system was to be a group of faculty associates who were willing (and qualified) to discuss improvement of teaching with their colleagues.

This system has been an enormous success--if we measure success by the amount of use. IDQ was used in 300 classes in the first year it was available. It was used in 1000 classes during its second year, in 2000 classes during its third year, in nearly 4000 classes during its third year, and in 5000 classes during its fifth year. Instructor-Designed Questionnaires are currently used in 7000 classes per year at Michigan.

In an important sense, however, IDQ did not develop as we intended it to. We originally designed the system to help individual faculty members who wanted to improve their teaching. Some faculty members do use IDQ primarily for self-improvement. They begin the evaluation process with self-reflection, and they complete it by consulting with fellow teachers. Those who use the system in this way usually report that it is useful, but such faculty members are a small minority of current users of IDQ.

Current Use

The real forces behind the growth of IDQ--and the forces that sustain its current level of activity--are the administration and the students at Michigan. With fewer and

fewer faculty slots available, Michigan administrators are emphasizing more than ever before that it is not enough for faculty members to be excellent researchers. They must also be excellent teachers, and there must be evidence that this is the case. Administrative needs have therefore led deans, department chairs, and teaching committees to order IDQ forms for all the faculty members under their jurisdiction. Forms ordered in this way are usually not as individualized as we would like them to be.

The second force that has contributed to the growth of IDQ at Michigan has been student consumerism. At the time that we began developing IDQ, costs of education were rising, and the potential economic payoff of education was declining. And students everywhere were beginning to ask how much they were getting for their educational dollar. Major student groups on campus began to support department-wide uses of IDQ, and they encouraged faculty members and departments to release results on certain items for publication in a booklet distributed to students. Numerous departments at the University agreed to cooperate and to release course evaluation data, and that added considerably to the growth of our course evaluation system. But it too cut down on the amount of individualization in forms.

The system that now operates at Michigan is different therefore from the system that we originally designed. It is now as much an external monitoring system as a self-monitoring program. And much of the praise--or blame--for the changes in the system has to be given to department administrators and student government groups on campus whose needs we eventually tried to meet through IDQ.

Future Directions

I think that further changes will occur in IDQ in the years ahead. One factor that may influence the direction in which the system moves is the growing perception at the University that statistical analyses of aggregate IDQ results can provide useful data for program evaluation. Teaching committees at the University now request such analyses of IDQ results fairly often, and we have seen increasing evidence that they use the results in program planning. In the time remaining this afternoon, I will describe one recent request for a special analysis of IDQ results.

The request came from the University's English Language Institute. This unit now has responsibility at the University for assessing communication ability of graduate students from non-English-speaking countries who are scheduled to be teaching assistants at the University. On the basis of observation of these foreign TAs, the English

Language Institute makes recommendations about their placement or non-placement in University of Michigan classrooms. During the past year, the English Language Institute approached CRLT to find out how TAs fare as teachers at the University.

We felt that the English Language Institute was focussing on an issue--the performance of teaching assistants--that was an important one in higher education generally. And the issue seemed to be an especially important one at Michigan--where close to half the credit hours earned by undergraduates are received in classes taught by teaching assistants and where an increasing number of teaching assistants are from non-English-speaking countries.

To help the English Language Institute get data on this topic, we focussed on four departments with significant numbers of new foreign teaching assistants: chemistry, economics, mathematics, and political science. We looked at student ratings of newly appointed foreign TAs in these departments, and we also looked at reactions to newly appointed TAs who were native speakers of English. Our analysis covered both the first and second semesters of teaching by these graduate students.

Before describing our findings on foreign teaching assistants, let me describe the picture that emerged of TAs for whom English was the native language (Table 1). During their first semester of teaching, native-language TAs fell far below the University mean on the items that we examined. On a typical item on overall excellence of teaching, instructional clarity, or teacher-student rapport, the native-language TAs got ratings about 0.8 standard deviations below the mean on University of Michigan norms. That means that the native-language TAs were at about the 25th percentile by UM standards. These low ratings did not seem to be a function of the level, size, or content of the classes taught by TAs. On an item asking students about their initial expectations, ratings for these classes were much nearer to the University average.

By the second semester, native-language TAs were performing approximately at the University of Michigan norm on the evaluation items that we examined (Table 2). Some of the improvement was due to self-selection. The better TAs continued teaching for a second-semester, whereas some of the TAs with lower ratings stopped teaching or found other teaching jobs. But improvement on the part of native-language TAs was another important factor in the rating changes. Native-language TAs who continued to teach improved significantly in ratings from the first to the second semester.

Results for foreign TAs were different from these results for native-language TAs. During the first semester, the foreign TAs were significantly lower than the native TAs in each of the areas that we examined. On excellence of teaching, instructional clarity, and student-teacher rapport, the scores of the foreign TAs were approximately 0.5 standard deviations below the scores of the native-language TAs. Scores for foreign TAs were thus 1.3 standard deviations lower than the all-University average. Typical scores for the foreign TAs were at the 13th percentile during their first semester of teaching.

Foreign TAs could be classified into two groups: those rated as proficient in English by the English Language Institute and those rated as non-proficient (Table 3). The proficiency ratings predicted reasonably well to student ratings of these teachers. Those foreign TAs who were rated as proficient got higher ratings from students than did those foreign TAs who were rated as non-proficient. Even those foreign TAs who were rated as proficient, however, received lower ratings than did native-language TAs. A program of selecting foreign TAs for classroom assignments on the basis of brief proficiency tests would thus alleviate somewhat the problem of poor student reactions to foreign TAs, but it would not entirely solve the problem.

The gap between foreign TAs and native ones does not close after one semester (Table 2). The foreign TAs who stopped teaching were not much different from those who continued in their classroom assignments, and the improvement in ratings of foreign TAs who continued teaching was not as great as it was for native TAs. As a result of these factors, the gap between ratings of foreign and native TAs was greater at the end of the second semester than it was at the end of the first semester.

Conclusions

What have we learned about course evaluation systems like IDQ from our experience at Michigan? First of all, I think that we have learned that such systems can provide useful information for teachers wishing to improve their teaching, but the proportion of teachers who will use these systems for this purpose without outside pressures is probably small. Second, we have learned that these systems can provide useful data for administrators and committees deliberating about personnel. And third, we have learned that these systems can provide useful data for students choosing between classes, and that such students will become strong advocates of these systems.

What I have emphasized today, however, is none of these major lessons from the use of IDQ. I have focussed instead

on an emerging new use of our data set in program evaluation. I have given only one example, but I hope that it gives a flavor of the sort of analyses that departments have begun to request at Michigan. Departments are asking us to carry out analyses that pull together data from different years, different areas, and that merge different data sets. They hope in this way to get insights into factors that influence their success with students so that policies and practices can be revised to increase their chances of success.

I believe that the University's program for training TAs will ultimately benefit from these types of analyses. The English Language Institute now has a better idea of the degree of success it achieved in its initial proficiency testing program, and baseline data now exist for the University to use in assessing a new program that is being designed for training foreign TAs.

Equally important, I think that we learned some lessons of general importance from helping the English Language Institute assemble and analyze these data. First, we learned that student ratings can change dramatically after an individual's first semester of teaching. I think that students and administrators should take this into account when they review teacher ratings. I think that perhaps a teacher's first semester ratings should be for the teacher's eyes only because first-semester ratings are open to misinterpretation. Second, verbal facility is an enormously important factor in teaching. Those who are linguistically handicapped seemed to be handicapped as teachers. Although it seems likely that linguistic deficiencies can be overcome, it is not yet clear how much time and effort must be spent by teachers who want to overcome such handicaps.

Table 1

First Semester IDQ Ratings of Native-language and Second-language Teaching Assistants

IDQ Item	Native-language TAs (N=46)			Second-language TAs (N=26)		
	M	z-score	%ile	M	z-score	%ile
Overall instructor rating						
Instructor is excellent.	3.52	-0.80	21	3.13	-1.35	13
Instructor is motivating.	3.23	-0.72	29	2.93	-1.25	15
Course is excellent.	3.42	-0.90	22	3.13	-1.45	13
Clarity of instructor						
Material presented clearly.	3.52	-0.75	21	3.02	-1.38	12
Requirements clearly defined.	3.87	-0.82	30	3.74	-1.28	17
Instructor-student rapport						
Comfortable asking questions.	4.09	-0.36	40	3.61	-1.29	13
Student difficulties recognized.	3.41	-0.76	29	3.08	-1.55	14

NOTE: The %ile rank is based on the distribution of class ratings from 2358 classes rated in fall 1983. The z-score gives the number of standard-deviation units separating the mean of the group from the all-University average on a given item.

Table 2

First Semester IDQ Ratings of Native-language Teaching Assistants and Second-language Teaching Assistants Classified as Proficient and Nonproficient at ELI Testing

IDQ Item	Native-language TAs (N=46)			Second-language TAs					
				Proficient (N=15)			Nonproficient (N=11)		
	M	z-score	%ile rank	M	z-score	%ile rank	M	z-score	%ile rank
Overall instructor rating									
Instructor is excellent.	3.52	-0.80	21	3.28	-1.14	17	2.94	-1.63	9
Instructor is motivating.	3.23	-0.72	29	2.99	-1.13	20	2.84	-1.41	11
Course is excellent.	3.42	-0.90	22	3.30	-1.12	19	2.90	-1.88	8
Clarity of instructor									
Material presented clearly.	3.52	-0.75	21	3.35	-0.96	19	2.54	-2.00	6
Requirements clearly defined.	3.87	-0.82	30	3.79	-1.09	22	3.68	-1.51	17
Instructor-student rapport									
Comfortable asking questions.	4.09	-0.36	40	3.83	-0.86	20	3.35	-1.80	7
Student difficulties recognized.	3.41	-0.76	29	3.24	-1.16	18	2.90	-2.00	7

NOTE: The %ile rank is based on the distribution of class ratings from 2358 classes rated in fall 1983. The z-score gives the number of standard-deviation units separating the mean of the group from the all-University average on a given item.

Table 3

Second Semester IDQ Ratings of Native-language and Second-language Teaching Assistants

IDQ Item	Native-language TAs (N=25)			Second-language TAs (N=20)		
	M	z-score	%ile rank	M	z-score	%ile rank
Overall instructor rating						
Instructor is excellent.	4.10	.01	54	3.25	-1.18	15
Instructor is motivating.	3.70	.12	57	3.02	-1.09	20
Course is excellent.	3.77	-.24	46	3.30	-1.14	19
Clarity of instructor						
Material presented clearly.	4.08	-.03	55	3.29	-1.04	17
Requirements clearly defined.	4.07	-.10	56	3.79	-1.11	22
Instructor-student rapport						
Comfortable asking questions.	4.34	.14	56	3.71	-1.10	16
Student difficulties recognized.	3.73	.03	53	3.18	-1.32	18

NOTE: The %ile rank is based on the distribution of class ratings from 2358 classes rated in fall 1983. The z-score gives the number of standard-deviation units separating the mean of the group from the all-University average on a given item.